Citation for published version:
Puttick, M, Guillerme, T & Wills, M 2020, 'The complex effects of mass extinctions on morphological disparity', *Evolution*, vol. 74, no. 10, pp. 2207-2220. https://doi.org/10.1111/evo.14078

DOI:
10.1111/evo.14078

Publication date:
2020

Document Version
Peer reviewed version

Link to publication

This is the peer reviewed version of the following article: Puttick, M.N., Guillerme, T. and Wills, M.A. (2020), The complex effects of mass extinctions on morphological disparity. Evolution, which has been published in final form at https://doi.org/10.1111/evo.14078. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. May. 2023

TITLE: The complex effects of mass extinctions on morphological disparity

Short title: Mass extinctions and morphological disparity

Mark N Puttick, Thomas Guillerme, Matthew A Wills

ABSTRACT

Studies of biodiversity through deep time have been a staple for biologists and palaeontologists for over 60 years. Investigations of species richness (diversity) revealed that at least five mass extinctions punctuated the last half billion years, each seeing the rapid demise of a large proportion of contemporary taxa. In contrast to diversity, the response of morphological diversity (disparity) to mass extinctions is unclear. Generally, diversity and disparity are decoupled, such that diversity may decline as morphological disparity increases, and vice versa. Here, we develop simulations to model disparity changes across mass extinctions using continuous traits and birth-death trees. We find no simple null for disparity change following a mass extinction but do observe general patterns. The range of trait values decreases following either random or trait-selective mass extinctions, whereas variance and the density of morphospace occupation only decline following trait-selective events. General trends may differentiate random and trait-selective mass extinctions, but methods struggle to identify trait selectivity. Long-term effects of mass extinction trait selectivity change support for phylogenetic comparative methods away from the simulated Brownian motion towards Ornstein-Uhlenbeck and Early Burst models. We find that morphological change over mass extinction is best studied by quantifying multiple aspects of morphospace occupation.

Keywords: disparity, mass extinctions, phylogenetic comparative methods, traits, macroevolution

INTRODUCTION

Evolutionary biologists and paleobiologists have long quantified diversity in terms of species numbers or species richness, making comparisons both horizontally between clades and higher taxa (Wiens 2017) and vertically throughout evolutionary time (Smith 2007). However, diversity takes no account of the morphological differences between species, a property known as morphological disparity (Wills et al. 1994). Researchers have attempted to formally define disparity in different ways, but most of the indices that derive from these definitions quantify variation in morphology or phenotype (Wills 2001; Hopkins and Gerber 2017; Guillerme et al. 2020a).

There have been numerous empirical analyses of the diversity of the global biota through evolutionary time (Raup and Sepkoski 1982; Alroy 2010) with an overall null expectation of symmetry as clades rise and fall in diversity (Gilinsky and Good 1989; Liow et al. 2010). For disparity, by contrast, there is no such simple null (Pie and Weitz 2005; Hughes et al. 2013; Oyston et al. 2015, 2016) but see (Foote 1996). Quantifying disparity alongside diversity is essential to fully understand the evolution of biodiversity (Roy and Foote 1997) as a series of empirical studies have demonstrated that diversity and disparity are largely decoupled (Wills et al. 1994; Fortey et al. 1996; Bapst et al. 2012). Analyses of disparity have accordingly proven invaluable for studying the tempo and mode of evolution (Simpson 1944), how clades diversify through time and throughout morphological "form space" (Gould 1990), and the patterns of drift and selection that have produced the distribution of living diversity (Raup 1981).

Extinction has had an immediate and potentially catastrophic role in sculpting patterns of biodiversity through time (Sepkoski 1981), particularly mass extinctions (Bond and Grasby 2017). Mass extinctions are defined, by convention, as geologically brief events that remove at least 75% of contemporary diversity (Barnosky et al. 2011). Researchers accept

45 that the effect of a mass extinction upon disparity will depend upon whether species are
46 selectivity or randomly removed by extinction (Foote 1997; Korn et al. 2013). Discrete
47 events in which extinction acts selectively are expected to decrease disparity by eliminating
48 the majority of an enclosed area of morphospace. Non-selective events, by contrast, may not
49 lead to disparity decreases (Foote 1991, 1993; Roy and Foote 1997; Villier and Korn 2004;
50 Korn et al. 2013), particularly when disparity is measured using a variance-based index (Korn
51 et al. 2013).

52 Empirical studies have demonstrated cases where mass extinctions acted both
53 selectively and non-selectively with regards to particular traits (Foote 1993; Roy 1996;
54 Lockwood 2004; Erwin 2007; Halliday and Goswami 2016). Phylogenetic comparative
55 studies indicate that there is trait selectivity on some groups, such as vascular plants (Green et
56 al. 2011), but there is little evidence to link traits and extinction susceptibility during mass
57 crises in the fossil record (Friedman 2009; Puttick et al. 2017; Allen et al. 2019). Many
58 studies reveal a phylogenetic signal of extinction (Hardy et al. 2012; Harnik et al. 2014; Krug
59 and Patzkowsky 2015; Puttick et al. 2017; Soul and Friedman 2017), without demonstrating
60 links to trait selectivity. We note that in studies of fossil record disparity, it is only possible to
61 analyze traits with fossilization potential, which excludes soft body parts and behavior.

62 In order to investigate the consequences of extinctions on traits, one approach would
63 be to analyze empirical data, as simulations may lack biological realism. Importantly,
64 however, simulations provide an underpinning framework for such empirical analyses (Foote
65 1991, 1997; Ciampaglio et al. 2001; Bapst et al. 2012; Korn et al. 2013), as it is possible to
66 test scenarios in which we know the definitive underlying patterns; an impossibility with
67 empirical data. Here we use a novel simulation approach to determine the expected, null
68 patterns of disparity change through mass extinction events and to investigate how quickly
69 diversity and disparity might be expected to recover to pre-extinction levels under many

scenarios. We simulate birth-death trees (Ma 2010; Stadler 2010; Mooers et al. 2012) and traits under Brownian motion with a variety of parameter values. During our simulations, a mass extinction removes a proportion of contemporary diversity. We show that disparity generally reduces following a mass extinction event, but that all patterns are variable. Alternative disparity indices imply different patterns of disparity change, such that there is a fundamental link between how we conceptualize and quantify disparity (Wills 1998b; Korn et al. 2013). Finally, we assess how mass extinctions bias comparative phylogenetic models of trait evolution in the extant lineages that survive such events. Phylogenetic regressions do not distinguish selective and non-selective mass extinction events under the simplest simulation models and the selective removal of taxa according to trait value causes comparative models to support non-BM models in preference to the (true) simulated BM process.

METHODS

Overview

We summarize the full simulation procedure in Fig 1. We simulated birth-death trees and Brownian motion traits simultaneously. We ran simulations until 50, 100, or 200 contemporary lineages were present, after which an extinction event removed 50, 75, or 90% of standing diversity. Simulations finished when diversity had recovered to pre-extinction levels (i.e., 50, 100, 200 tips). We chose the severity of mass extinctions (50, 75, 90%) to reflect known extinction values in the fossil record (Stanley 2016; Bond and Grasby 2017). The upper figure (90%) reflects the severity often quoted for the end-Permian event, although this has been questioned (Stanley 2016). The middle figure (75% of species lost) reflects the severest estimates for end-Ordovician and end-Cretaceous events and is accepted as the necessary threshold for a mass extinction (Barnosky et al. 2011). We used 50% for a

95 ‘smaller’ extinction events, such as the Frasnian extinction in the Late Devonian (Stanley
96 2016; Bond and Grasby 2017).

97 We simulated extinction in two different ways. Taxa were either eliminated at random
98 in a non-selective event or removed according to their trait values to simulate a selective
99 extinction. In separate analyses, we ran models to simulate one, two, and five traits. We then
100 assessed disparity through time and analyzed the ability of models to recover the signal of
101 trait change across extinctions. We wrote new code in *R* to simulate data; users can freely
102 download this as part of the package *Phylogenetic Evolution of Traits and Extinction in R*
103 (PETER) from *GitHub* (https://github.com/PuttickMacroevolution/PETER).

104 The first objective of this study is to provide a simple null model for change in
105 disparity across mass extinctions, given a number of varying parameters. In our simulations,
106 we have an extremely high resolution of extinction through time as we record the trait values
107 for all simulated internal nodes and tips. In the fossil record, however, it is often necessary to
108 aggregate data into time bins. Therefore, we also undertook steps to study binned values of
109 disparity through time. For the majority of our analyses, we focused on studies with one
110 continuous trait, as there is often just a single measurement available for a large group of
111 species (e.g., body size in fossil vertebrates). For these simulations, we attempted to control
112 for properties that are likely to vary across the tree of life, such as rates of background
113 extinction and the total number of species. Our multivariate approach was designed to extend
114 the single trait approach, albeit that selectivity was simulated as acting on a single trait.
115 Evolution in this directly selected focal trait was correlated with evolution in the other
116 trait(s), so that these latter traits experienced indirect selectivity (Lande and Arnold 1983).
117 We also ran simulations in which traits evolved independently.

118 The second objective is to examine the effectiveness of methods commonly used to
119 investigate mass extinctions. For example, we determine whether it is possible to differentiate

120 trait-selective and trait-independent mass extinctions, and how mass extinctions influence
121 models of continuous trait evolution through time.

122 ***Simulation of trees and traits***

123 We simulated trees evolving from the root to the tips with time-homogeneous
124 speciation (λ) and extinction rates (μ). We used the same speciation rate ($\lambda = 1$) for all
125 simulations, but different levels of background extinction ($\mu = 0, 0.4, 0.8$).

126 We simulated traits under a Brownian motion process (Felsenstein 1973, 1985). At
127 the start of each simulation, we sampled the root state from a normal distribution with a mean
128 of zero and a standard deviation of one. In simulations with two and five traits, we set the co-
129 variance either to zero such that the traits were independent or to a co-variance of 0.75 so that
130 the evolution of traits was linked but not identical. The use of a moderate co-variance (0.75)
131 implies that although selectivity (when present) would act on one trait only, the mass
132 extinction would also affect the remaining, non-directly selected traits. We designed this to
133 mimic a scenario in which selectivity acts on one analyzed trait that is known to co-vary with
134 many other traits, such as body size.

135 ***Disparity indices***

136 We summarized diversity using lineage-through-time plots, and we used three indices
137 of disparity (Fig 1C): *median pairwise distance*, which is the median pairwise Euclidean
138 distance between all points (Wills et al. 1994); *Sum of Variances* (SOV) from each trait
139 (Foote 1992a); and *Sum of Ranges* (SOR) (Wills 1998a, 2001; Brusatte et al. 2011; Ruta et al.
140 2013). We use these disparity indices to elucidate different aspects of morphospace
141 occupation: we interpret *median pairwise distance* as an index of the density of morphospace
142 occupation; and we consider *Sum Of Variance*s as an index of dispersion, and *Sum Of Ranges*
143 as an index of the overall magnitude or volume of morphospace occupation (Guillerme et al.

144 2020b). Unless stated, all analyses were conducted using the R package *dispRity* (Guillerme
145 2018).

146 Our simulations produced trait estimates for every node in the tree. We estimated
147 disparity indices using these traits, but for most analyses, we binned trait values into equal-
148 size time bins and subsequently calculated disparity for each bin using a time-slicing
149 approach. We explored the impacts of using different bin sizes (4, 8, 10, 16-time bins), and
150 used the highest resolution, sixteen-time bins to assess if there were significant differences in
151 using each bin size. We used sixteen time-bins throughout the remaining analyses, but we
152 observed the same general trends with all bin sizes (Table 1). After each simulation, we
153 scaled disparity and diversity to unity by dividing each by their maximum values.

154 **Simulation of mass extinctions**

155 For a given quantile of extinction, we set different strengths of selectivity:

156 (i) *Strict selectivity*. All lineages with traits larger than the cut-off went extinct and all
157 other lineages survived.

158 (ii) *Strong selectivity*. To investigate events in which the vast majority of extinction
159 related to the size of traits, lineages larger than the cut off had a probability of
160 0.99 to go extinct whilst lineages smaller than the cut off had a probability of 0.01
161 to go extinct.

162 (iii) *Random selectivity*. Trait values had no influence on extinction susceptibility.

163 (iv) *Weak selectivity.* An intermediate scenario in which lineages larger than the cut off
164 had a probability of 0.75 of extinction whilst lineages smaller than the cut off had
165 an extinction probability of 0.25 (i.e, there was a 25% chance of random
166 extinction).

167 In each case of selectivity strength, we set the cut-off by using the contemporary
168 distribution of traits at the extinction boundary; lineages with trait values above the 0.5, 0.25,

or 0.1 quantiles of the contemporary distribution were more prone to extinction in those simulations with *directional selectivity*. The quantiles were chosen to directly reflect the selectivity of extinction. So, for simulations with *strict selectivity* there was a zero probability of extinction for taxa with trait values below a certain cut-off value. As an example, for simulations with 75% loss of taxa at an extinction boundary and *strict selectivity*, there was a zero probability of extinction for taxa with trait values below the 0.25 quantile of all contemporary trait values.

We binned and summarized simulated data by pooling all parameters: *n* lineages lost, background extinction, and the number of tips for *strict*, *strong*, and *random* extinctions. We used three levels of background extinction in all other simulations (i.e., 0, 0.4, 0.8). In analyses with one trait only, we also ran simulations with a background extinction rate of 1.

Finally, we ran *disruptive selectivity* mass extinction simulations for a single trait in which lineages with trait values closer to the mean were prone to extinction, such that lineages with smaller or larger trait values were more likely to survive. For these simulations, we set the selective extinctions to remove lineages from the 0.5, 0.75, and 0.9 quantiles symmetrical around the mean.

Phylogenetic analysis of mass extinctions

We tested whether survivors and extinct lineages differed significantly in their trait values using Ordinary Least Squares (OLS) and Phylogenetic Generalized Least Squares (pGLS) regressions (Grafen 1989). In these models, we set the response as the continuous trait value and coded the predictor as a binary variable indicating whether or not the lineage survived the mass extinction. We pruned the phylogeny and trait values to include only those lineages that existed at the boundary (Puttick et al. 2017), such that we did not consider values from lineages that were lost before the extinction or arose after it. We estimated the

phylogenetic signal in trait values using Pagel's lambda (Pagel 1997, 1999) alongside the regression parameters in the R package *motmot* (Puttick et al. 2020).

We tested the phylogenetic signal in extinction itself, without consideration of trait values, using the *phyloD* statistic (Fritz and Purvis 2010). This method used phylogenetic contrasts to estimate internal node values of binary traits (zero = survivors, one = extinct lineages), and used these node values to estimate the number of binary transitions. The number of changes indicated the phylogenetic signal, with the number of changes and signal strength being inversely correlated (more changes means a weaker phylogenetic signal). A Brownian motion model will produce a *d* value close to 0 under the *phyloD* statistic. Values larger than 0 denote increasingly random signal, and values under 0 are indicative of over-conserved signals.

Phylogenetic comparative models of trait distributions that have recovered from a mass extinction

The signal of past mass extinctions may be present in the trait distribution of extant lineages. To explore these patterns, we pruned simulated datasets to yield ultrametric trees comprising only extant taxa (i.e., lineages in existence at the end of the simulation). We applied commonly-used likelihood phylogenetic comparative models of trait evolution to these pruned data: Brownian motion (Felsenstein 1973, 1985) to represent the null model; evolution under a constraint using an Ornstein-Uhlenbeck (OU) model (Hansen 1997; Butler and King 2004; Blomberg et al. 2020) in which variance was constrained to the ancestral value according to the strength of α; and the Early Burst model, in which the rate variance exponentially decreased through time (Blomberg et al. 2003; Harmon et al. 2010).

The whole procedure for running these simulations is available and documented on *GitHub* (https://github.com/PuttickMacroevolution/PETER).

RESULTS

The variance around simulated disparity is higher than the variance around diversity

We ran initial analyses without a mass extinction event. When plotted through time, relative disparity showed an overall pattern of increase (Fig 2). Trends are similar for each index, but there is a more substantial difference between disparity and diversity when the former is measured using the Sum Of Ranges rather than the Sum of Variances or median pairwise distance. We observed similar trends in simulations with two and five traits (Supplementary Fig S1).

Mass extinctions usually cause reductions in disparity, and the precise patterns are related to the index of disparity

As a general rule, selective mass extinctions caused a reduction in disparity (Fig 3, Supplementary Figure S2). However, this pattern was not universal, and changed according to the simulation model (e.g., different strengths of selectivity) and the disparity index used (Fig 4).

All disparity indices showed iterations with both increases and decreases across a mass extinction, such that there is no universal pattern. Pooling the results of all simulations that entailed strict directional selectivity (irrespective of background extinction, number of tips, or extinction severity) we found a median decrease in disparity indexed by median pairwise distance (90%), Sum of Variances (90%), and Sum of Ranges (75%). When selectivity was absent from the simulations, all indices showed a median disparity change of zero across the mass extinction, except for the Sum Of Ranges for which most simulations showed a decrease in disparity across the boundary (75%). For those simulations with one trait, we ran additional simulations that used the special case of $\lambda = \mu = 1$, and these showed patterns similar to those in simulations with other levels of extinction (Supplementary Table S1, Supplementary Figure S2).

For a single trait, we also investigated the effects of varying the simulation parameters individually (viz., strength of selectivity, the proportion of lineages lost, rate of background extinction, number of tips, and disparity index used) (Supplementary Fig S3). Simulations with more tips showed more extreme reductions in disparity across the mass extinction boundary, irrespective of the disparity index used. Only in certain circumstances did all iterations indicate a decrease in disparity across an extinction boundary. For example, there was a consistent decrease in disparity across a mass extinction in simulations with more tips, more severe lineage loss, and higher rates of background extinction (including a background extinction of $\mu = 1$). The Sum Of Ranges index is very sensitive to sample size differences, and a fall in the Sum of Ranges would be expected alongside a reduction in diversity in the wake of a mass extinction. For the *directional selectivity* simulations with one trait, we ran additional simulations that used the special case of $\lambda = \mu = 1$ and these showed similar patterns to simulations with other levels of extinction (Supplementary Table S1).

For simulations with an intermediate level of extinction selectivity (*weak selectivity* with 25% probability of random extinction), the results are generally closer to *random selectivity* models rather than to other models with directional selectivity (Supplementary Table S1).

In simulations with multiple traits, only the extinctions with strict selectivity showed a consistent decrease in disparity across a mass extinction boundary (Fig 4B-C). The median decreases in disparity for multiple traits were smaller than in simulations with one trait, and some multiple trait simulations indicated an increase in disparity across the extinction event. The number of iterations that showed a decrease in disparity across the mass extinction boundary was lower in simulations with no trait co-variance compared to analyses with 0.75 trait co-variance (Table 1A). Again, an exception was the Sum Of Ranges, in which the majority of iterations showed a decrease in disparity with any selectivity strength, and nearly

all traits had a decrease in disparity with five traits (Table 1B). We note that increases in range-based indices are only possible because of time binning effects: the *immediate* wake of an extinction event (if not time averaged and therefore excluding new linages) could only show a decrease or no change in range. However, we seek to simulate the temporal sampling that pertains in typical paleontogical data sets. The tendency for a disruptive extinction was for all disparity indices to show an increase or no change across the mass extinction boundary (Fig S4).

Comparative phylogenetic methods do not distinguish selective and non-selective extinctions

pGLS models were more conservative at detecting the true relationship between trait values and extinction selectivity compared to OLS models, but pGLS models have lower type-two error rates in comparison to OLS models (Fig 5). In all simulations with *strict* and *strong* selectivity 99% of OLS models support the correct relationship, compared to only a median of 86% (*strict*) and 68% (*strong*) of pGLS models. With *weak selectivity,* few pGLS (median 6%) models show a significant relationship between extinction and traits, which is much lower than the OLS relationship (80%). When there is no relationship between trait values and extinction selectivity, pGLS models have a lower error rate (median 1%) compared to pGLS models (median 6%).

Models consistently detect the phylogenetic signal of extinction

There was a strong phylogenetic signal, as measured by the *phyloD* statistic, when selectivity was *strict* (median 0.07) and *strong* (median 0.12). Conversely, there was a random signal when selectivity was *weak* (median 0.82) or *random* (median 1.0). Only the strength of selectivity and magnitude of lineage loss were significant factors in explaining the differences in phylogenetic signal. There was no significant difference in the phylogenetic signal for different numbers of tips or different levels of background extinction.

292 ***Mass extinctions have long-term impacts on trait distributions***

293 High levels of extinction selectivity increased support for non-Brownian models when
294 analyzed using exclusively extant data that had ancestrally suffered a mass extinction (Fig 6,
295 Supplementary Fig S5). For a single trait with no mass extinction, most simulations
296 supported a Brownian motion model (median 82%), with relatively few supporting Early
297 Burst (3%) and OU (15%) models.

298 For simulations of *directional selectivity*, strong or strict selectivity yielded much
299 lower support for Brownian motion (median strict 51%, strong 70%), and higher support for
300 the OU model (strict 49%, strong 28%). When there was no selectivity on the trait value at
301 the mass extinction boundary, the relative support for models was comparable to support for
302 models in simulations with no mass extinctions (BM support, no selectivity 80%). Across all
303 analyses that supported the OU model, the strength of selectivity α had a median half-life of
304 1.69 (within a range of values from 0.52 to 17.25). Similar patterns were seen in simulations
305 with two traits compared to simulations with one trait.

306 With higher levels of background selectivity in the simulations with *disruptive*
307 *selectivity* at mass extinctions, there was lower support for the OU model compared to BM.
308 This was because there was a high turnover of lineages such that the 'crown' of these
309 simulated phylogenies emerged after the mass extinction. When only those trees that had at
310 least one 'crown' node *predating* the mass extinction were analyzed, the support for OU
311 model was higher (Fig S6). *Disruptive selectivity,* in contrast to directional trait selectivity,
312 resulted in higher support for an early burst model compared with the simulated Brownian
313 motion process.

314 **DISCUSSION**

315 Few putative macroevolutionary rules withstand scrutiny (McShea 1998; Hone and
316 Benton 2005; McShea and Brandon 2010; Benson et al. 2018). However, there do appear to

317 be statistical generalities concerning the manner in which major clades evolve through time,
318 with most groups achieving maximum or near maximal morphological disparity relatively
319 early in their existence (Foote 1992a, 1994, 1996; Hughes et al. 2013; Oyston et al. 2015,
320 2016). The largest environmental crises often appear to disrupt this pattern, with those clades
321 that go extinct coincident with a mass event typically being truncated and having maximum
322 disparity much later in their evolutionary trajectories (Hughes et al. 2013). Despite these
323 empirical observations, there are no null models for the manner in which we expect the
324 disparity of clades to change through time, still less for clades truncated by or (as here)
325 surviving such events with reduced diversity (Korn et al. 2013). Hence, simulations provide a
326 powerful way of analyzing the role of mass extinctions in shaping disparity.
327 We acknowledge that there is no single, universally agreed index for disparity.
328 Moreover, all empirical assessments of disparity are necessarily relative and constrained
329 within the context of a particular set of descriptors or data set. We also note that disparity is
330 indexed with reference to the constituent entities (species or other operational taxonomic
331 units) within a group (e.g., subclade or time bin), with no reference to species or other entities
332 outside of that group. As such, identical indices of disparity can be reported for clades
333 occupying very different regions of morphospace, or for clades that migrate through that
334 morphospace through time. Strikingly different distributions of points (local densities and
335 clustering structures) can also yield identical disparity indices (Wills et al. 2012).
336 Comparisons of diversity and disparity can nevertheless yield insights into the dynamics of
337 evolutionary change (Foote 1991, 1993, 1994, 1997), while comparisons of a variety of
338 disparity indices can be used to classify patterns of disparity change across mass extinction
339 boundaries (Korn et al. 2013). There are many aspects of evolutionary dynamics (e.g., the
340 directionality or otherwise of selectivity, modes of morphological evolution or speciation)
341 that are more effectively tested and modelled directly rather than via the distributions of taxa

342 in morphospace.

343 By definition, a mass extinction destroys standing diversity, but our results suggest
344 that it need not always precipitate a fall in disparity (Foote 1991). Across a selective
345 extinction the variance and range of morphological occupation typically decreases (Fig. 3);
346 but the variance is mostly unchanged across a non-selective extinction, even though the range
347 of values still decreases. The patterns of disparity change that we observe depend upon the
348 indices that we use to quantify it, so these indices must therefore be codifying different
349 things. Only the simulations with little deviation from trait-based extinctions show general,
350 predictable patterns of extinction through time. However, when there is more variance around
351 mass extinction selectivity (*weak selectivity*), the observed patterns are much closer to those
352 seen when extinction is random in regard to trait values.

353 Patterns of disparity are more complicated than patterns of diversity. Here, we
354 principally discuss patterns of directionally selective mass extinctions, but we note that
355 disruptive selectivity can yield disparity increases across extinction boundaries (Fig S4).
356 Many parameters complicate the pattern of disparity change across mass extinction events,
357 including the number of analyzed traits, co-variance between traits, and the index of disparity
358 (Fig 3; Fig S3). For example, when the co-variation between multiple traits is high (0.75), the
359 observed patterns of disparity through time (Fig 4B) are similar to those observed with just
360 one trait (Fig 4A). Thus, there is no single expectation of disparity change across a mass
361 extinction for all clades and all events. Rather, we suggest that in order to understand patterns
362 of disparity change across a mass extinction for a particular group, it is necessary to
363 determine whether this pattern differs from those expected given a similar set of parameters.
364 These expected patterns could be determined using simulations (Foote 1991; Harmon et al.
365 2003; Slater et al. 2010; Green et al. 2011; Korn et al. 2013). We do note, however, that even

366 with simulation approaches it could be impossible to detect small-scale changes in selectivity
367 at a large scale (Raup et al. 1973), especially in analyses with hundreds or thousands of traits.
368 Where we find decreases in disparity across mass extinction events, these are often
369 non-linear (Fig 3). Clades evolve through an empirically realized (Stone 1997) or
370 theoretically possible (Novack-Gottshall 2007) multidimensional trait morphospace via a
371 process of branching cladogenesis that is inherently diffusive and with increasing degrees of
372 freedom (McShea 1998). By contrast, random processes of lineage extinction cause a linear
373 decline in diversity, but disparity is a function of the *distribution* of a group of entities, be this
374 morphological range, variance or otherwise (Foote and Gould 1992; Wills et al. 1994; Wills
375 2001; Ruta et al. 2013; Bazzi et al. 2018). Hence after an extinction event, disparity is
376 expected to decline non-linearly and more slowly than diversity (Foote 1991; Ruta et al.
377 2013). This expectation is generally echoed in our simulations (Fig 3), but the precise pattern
378 is contingent on the index of disparity used.
379 Korn et al. (2013) demonstrated that it is possible to differentiate types of mass
380 extinction selectivity by examining changes in morphospace occupation. For example, a
381 random extinction that is non-selective with regards to lineage trait values leads to a minor
382 decrease in the Sum Of Ranges, with no appreciable change in disparity indexed using the
383 Sum Of Variances. For median pairwise distance and Sum Of Variances indices, a substantial
384 decrease in disparity may be indicative of a directionally selective mass extinction. In some
385 instances, these indices may not reflect a decrease in disparity or may reflect a change in the
386 pattern of morphospace occupation (Ciampaglio et al. 2001). The Sum Of Ranges decreases
387 with all types of extinction selectivity and so does not differentiate between selective and
388 random extinctions as well as mean pairwise distance and Sum Of Variances. Because the
389 Sum Of Ranges is sensitive to sample size effects (Foote 1991, 1992b; Butler et al. 2012) it
390 may be an unsuitable index to capture patterns of disparity across a mass extinction event.

391 However, some have argued this is unimportant when sample sizes are large (Simon et al.
392 2010), and that it can be adjusted using rarefaction approaches (Foote 1992b; Wills 1998a).
393 Median pairwise distances and Sum Of Variances may be able to differentiate selective and
394 non-selective extinction, but no one index of disparity will be sufficient to distinguish all
395 aspects of evolutionary change (Ciampaglio et al. 2001). Moreover, we note that in studies
396 with a higher number of traits than the five simulated here, it will be even more challenging
397 to abstract generalities.

398 We have primarily investigated the patterns of a directionally selective extinction
399 acting on one trait. Further studies could examine reductions in multi-variate shape space at
400 extinction boundaries (Korn et al. 2013), by employing simulations of non-homogeneous
401 patterns of tree and trait evolution, and patterns of discrete character trait change. Trait
402 evolution in our models is simulated under Brownian motion, such that disparity levels tend
403 to increase linearly through time, and, in the absence of extinction, disparity is higher than
404 relative diversity in the early history of a clade (Figs 2). In the absence of selective extinction
405 (Korn et al. 2013), clades are expected to have top-heavy disparity profiles in which disparity
406 is higher later in time (Gould 1990; Hughes et al. 2013; Deline et al. 2018; Hill et al. 2018).
407 This observations contrasts strikingly with the empirical observation that there is a significant
408 but weak tendency for clades to have low diversity but high disparity relatively *early* in their
409 evolution (Foote 1994; Hughes et al. 2013). In our simulations we employ homogeneous
410 models of speciation and extinction, but alternative patterns are expected to arise from
411 heterogeneous diversification rates through time. Additionally, simulations and evolutionary
412 models can employ different relations between traits and extinction, such as modeling
413 extinction as a logistic function of trait values (Slater et al. 2017).

414 Rapid speciation early in a clade's history is expected to result in disparity being
415 higher between (rather than within) clades: this leads to a pattern in which average subclade

disparity decreases towards the present (Harmon et al. 2003). However, the patterns we present are not directly comparable to Disparity-Through-Time (DTT) approaches as DTT methods calculate disparity at each point in time using all tip descendants of nodes in a phylogenetic tree (Harmon et al. 2003). Here we estimate disparity based on all of the branches present at a point in time, or we otherwise pool these samples into time bins.

A key question in paleobiology concerns whether mass extinctions selectively remove taxa based on their trait values or whether extinction is random with regards to traits. To this end, previous phylogenetic approaches to detecting extinction selectivity at mass extinctions have utilized pGLS models to test whether surviving and extinct lineages differ significantly in trait values (Friedman 2009; Puttick et al. 2017; Allen et al. 2019). pGLS models are appropriate in this context to correct for non-independence of residuals in the regression as a result of shared phylogenetic history (Felsenstein 1985), but we show that pGLS models are not able to correctly support the hypothesis that trait values differ in extinct lineages compared to surviving lineage trait values (Fig 5). In contrast, non-phylogenetic Ordinary Least Squares analyses correctly support a significant difference between trait values in survivors and losers at a mass extinction boundary. The cost of Ordinary Least Squares approaches is that they have type-one error rates above 5%, so not using phylogenetic correction cannot be recommended (Felsenstein 1985). Thus, the lack of pGLS support for a selective extinction may not mean that the extinction was random but may reflect a type-two error. This may result from the strictly homogeneous pattern of evolution employed here, which may not reflect biological reality. However, the solution for this problem is unclear, as it is not straightforward to compare non-phylogenetic and phylogenetic modelling results (Freckleton 2009), and impossible to differentiate true signals from model error. One solution, albeit unsatisfactory, is to infer trait-selectivity by analyzing the phylogenetic signal in extinction, without considering traits directly. Previous research has indicated that it is

441 possible to differentiate selectivity types using the phylogenetic signal of extinction (Hardy et
442 al. 2012; Harnik et al. 2014; Krug and Patzkowsky 2015; Puttick et al. 2017; Soul and
443 Friedman 2017). We support these observations because selective and non-selective
444 extinctions have significantly different phylogenetic signals in our simulations (measured
445 using the *phyloD* statistic), and because the amount of phylogenetic diversity lost during an
446 extinction event is dramatically different for random and selective events. We note, however,
447 that even low levels of random noise (25% extinction non-selective, *weak selectivity*)
448 produces results much closer to a fully random extinction simulation rather than one with
449 selectivity.

450 After recovery from mass extinctions, extant species trait distributions still carry the
451 signature of trait-selective extinction. This selectivity of extinction towards specific
452 morphologies leads to changes in the relative support of phylogenetic comparative methods
453 away from the simulated Brownian motion model to support an Ornstein-Uhlenbeck process
454 (Fig 6A,C). Selective mass extinctions remove the left hand of the trait distributions, so the
455 variance of tip trait values is lower than expected under a Brownian motion model; this trait
456 distribution then resembles the expected distribution under the Ornstein-Uhlenbeck process,
457 so this model is supported (Fig 6E). At face value, this may appear to be a bias that does not
458 support the simulation model (BM) over an erroneous alternative model (OU). If OU is
459 interpreted as a measure of stabilizing selection (Hansen 1997), then support for this model is
460 correct in the context of a selective mass extinction. Taxa in a selective extinction are
461 removed as they possessed trait values that made them prone to extinction. The OU α value
462 from the models here can be interpreted as measuring a discrete event of stabilizing selection
463 (Butler and King 2004) or as an evolutionary optimum (Hansen 1997) that confers survival to
464 a mass extinction event. In an analogous situation, a disruptive extinction event leads to a bi-

modal distribution in which a distribution resembles an Early Burst pattern over alternative models (Fig 6b).

Conclusions

Our results indicate that patterns of disparity change across a mass extinction boundary are more complex than patterns of diversity change. As would be expected, disparity generally decreases following a selective extinction, but most indices show both increases and decreases in disparity across boundaries.

Understanding whether mass extinctions are selective or random with regards to traits is a major open question in paleobiology. Our results suggest that differentiating selective and non-selective extinction is difficult with phylogenetic comparative methods, but it is easier to identify phylogenetic signal in extinction itself. Mass extinctions also have long-term impacts on trait distributions and comparative models of trait evolution, even when traits evolve by a simple, time-homogeneous process.

All conclusions here assume that disparity may represent a real biological signal or a phenomenological description of patterns of trait evolution. Whatever trait disparity is measuring, we suggest that multiple indices are necessary in order to capture all aspects of disparity change across an extinction boundary, and a number of parameters need to be considered when inferring extinction disparity patterns.

ACKNOWLEDGEMENTS

MNP acknowledges support from the Royal Commission for the Exhibition of 1851. MAW acknowledges support from the Natural Environment Council grant NE/K014951/1 and John Templeton Foundation grant 61408. TG was funded by the ARC DP170103227 and FT180100634 awarded to Vera Weisbecker and the NERC grant number NE/T000139/1 awarded to Gavin Thomas

REFERENCES

490 Allen, B. J., T. L. Stubbs, M. J. Benton, and M. N. Puttick. 2019. Archosauromorph
491 extinction selectivity during the Triassic-Jurassic mass extinction. Palaeontology
492 62:211–224.
493 Alroy, J. 2010. The shifting balance of diversity among major marine animal groups. Science
494 329:1191–1194.
495 Bapst, D. W., P. C. Bullock, M. J. Melchin, H. D. Sheets, and C. E. Mitchell. 2012.
496 Graptoloid diversity and disparity became decoupled during the Ordovician mass
497 extinction. Proc. Natl. Acad. Sci. U. S. A. 109:3428–3433.
498 Barnosky, A. D., N. Matzke, S. Tomiya, G. O. U. Wogan, B. Swartz, T. B. Quental, C.
499 Marshall, J. L. McGuire, E. L. Lindsey, K. C. Maguire, B. Mersey, and E. a Ferrer.
500 2011. Has the Earth's sixth mass extinction already arrived? Nature 471:51–57.
501 Bazzi, M., B. P. Kear, H. Blom, P. E. Ahlberg, and N. E. Campione. 2018. Static Dental
502 Disparity and Morphological Turnover in Sharks across the End-Cretaceous Mass
503 Extinction. Curr. Biol. 28:2607-2615.e3.
504 Benson, R. B. J., G. Hunt, M. T. Carrano, and N. Campione. 2018. Cope's rule and the
505 adaptive landscape of dinosaur body size evolution. Palaeontology 61:13–48.
506 Blomberg, S. P., T. G. Jr, and A. R. Ives. 2003. Testing for phylogenetic signal in
507 comparative data: behavorial traits more labile. Evolution 57:717–745.
508 Blomberg, S. P., S. I. Rathnayake, and C. M. Moreau. 2020. Beyond brownian motion and
509 the ornstein-uhlenbeck process: Stochastic diffusion models for the evolution of
510 quantitative characters. Am. Nat. 195:145–165.
511 Bond, D. P. G., and S. E. Grasby. 2017. On the causes of mass extinctions. Palaeogeogr.
512 Palaeoclimatol. Palaeoecol. 478:3–29.
513 Brusatte, S. L., S. Montanari, H. Yi, M. A. Norell, S. L. Brusatte, S. Montanari, H. Yi, and
514 M. A. Norell. 2011. Phylogenetic corrections for morphological disparity analysis: new

methodology and case studies. Paleobiology 37:1–22.

Butler, M. A., and A. A. King. 2004. Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. Am. Nat. 164:683–695.

Butler, R. J., S. L. Brusatte, B. Andres, and R. B. J. Benson. 2012. How do geological sampling biases affect studies of morphological evolution in deep time? A case study of pterosaur (Reptilia: Archosauria) disparity. Evolution 66:147–162.

Ciampaglio, C. N., M. Kemp, and D. W. Mcshea. 2001. Detecting changes in morphospace occupation patterns in the fossil record: Characterization and analysis of measures of disparity. Paleobiology 27:695–715.

Deline, B., J. M. Greenwood, J. W. Clark, M. N. Puttick, K. J. Peterson, and P. C. J. Donoghue. 2018. Evolution of metazoan morphological disparity. Proc. Natl. Acad. Sci. U. S. A. 201810575.

Erwin, D. H. 2007. Disparity: Morphological pattern and developmental context. Palaeontology 50:57–73.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25:471–492.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Foote, M. 1993. Discordance and concordance between morphological and taxonomic diversity. Paleobiology 19:185–204.

Foote, M. 1996. Models of morphological diversification. Pp. 62–86 *in* D. Jablonski and D. H. Erwin, eds. Evolutionary Paleobiology. University of Chicago Press, Chicago.

Foote, M. 1991. Morphological and taxonomic diversity in clade's history: the blastoid record and stochastic simulations. Contrib. From Museum Paleontol. 28:101–140.

Foote, M. 1994. Morphological disparity in Ordovician-Devonian crinoids and early saturation of morphological space. Paleobiology 20:320–344.

Foote, M. 1992a. Paleozoic record of morphological diversity in blastozoan echinoderms. Proc. Natl. Acad. Sci. U. S. A. 89:7325–7329.

Foote, M. 1992b. Rarefaction analysis of morphological and taxonomic diversity. Paleobiology 18:1–16.

Foote, M. 1997. The evolution of morphological diversity. Annu. Rev. Ecol. Syst. 129–52.

Foote, M., and S. J. Gould. 1992. Cambrian and recent morphological disparity. Science 258:1816–1818.

Fortey, R. A., D. E. G. Briggs, and M. A. Wills. 1996. The Cambrian evolutionary "explosion": decoupling cladogenesis from morphological disparity. Biol. J. Linn. Soc. 57:13–33.

Freckleton, R. P. 2009. The seven deadly sins of comparative analysis. J. Evol. Biol. 22:1367–1375.

Friedman, M. 2009. Ecomorphological selectivity among marine teleost fishes during the end-Cretaceous extinction. Proc. Natl. Acad. Sci. U. S. A. 106:5218–5223.

Fritz, S. A., and A. Purvis. 2010. Selectivity in Mammalian Extinction Risk and Threat Types: a New Measure of Phylogenetic Signal Strength in Binary Traits. Conserv. Biol. 24:1042–1051.

Gilinsky, N. L., and I. J. Good. 1989. Analysis of clade shape using queueing theory and the fast fourier transform. Paleobiology 15:321–333.

Gould, S. J. 1990. Wonderful life: the Burgess Shale and the nature of history. WW Norton and Company.

Grafen. 1989. The Phylogenetic Regression. Philos. T. R. Soc. B. 326:119–157.

Green, W. A., G. Hunt, S. L. Wing, and W. A. DiMichele. 2011. Does extinction wield an axe or pruning shears? How interactions between phylogeny and ecology affect patterns of extinction. Paleobiology 37:72–91.

565 Guillerme, T. 2018. dispRity : A modular R package for measuring disparity. Methods Ecol.
566 Evol. 1755–1763.
567 Guillerme, T., N. Cooper, S. L. Brusatte, K. E. Davis, A. L. Jackson, S. Gerber, A. Goswami,
568 K. Healy, M. J. Hopkins, M. E. H. Jones, G. T. Lloyd, J. E. O'Reilly, A. Pate, M. N.
569 Puttick, E. J. Rayfield, E. E. Saupe, E. Sherratt, G. J. Slater, V. Weisbecker, G. H.
570 Thomas, and P. C. J. Donoghue. 2020a. Disparities in the analysis of morphological
571 disparity. Biol. Lett. 16:20200199.
572 Guillerme, T., M. N. Puttick, A. Marcy, and V. Weisbecker. 2020b. Which disparity or
573 dissimilarity measurement best summarize occupancy in multidimensional spaces? Ecol.
574 Evol., doi: https://doi.org/10.1002/ece3.6452.
575 Halliday, T. J. D., and A. Goswami. 2016. Eutherian morphological disparity across the end-
576 Cretaceous mass extinction. Biol. J. Linn. Soc. 118:152–168.
577 Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation.
578 Evolution 51:1341–1351.
579 Hardy, C., E. Fara, R. Laffont, J. L. Dommergues, C. Meister, and P. Neige. 2012. Deep-time
580 phylogenetic clustering of extinctions in an evolutionarily dynamic clade (early jurassic
581 ammonites). PLoS One 7:1–7.
582 Harmon, L. J., J. B. Losos, T. Jonathan Davies, R. G. Gillespie, J. L. Gittleman, W. Bryan
583 Jennings, K. H. Kozak, M. A. McPeek, F. Moreno-Roark, T. J. Near, A. Purvis, R. E.
584 Ricklefs, D. Schluter, J. A. Schulte, O. Seehausen, B. L. Sidlauskas, O. Torres-Carvajal,
585 J. T. Weir, and A. T. Mooers. 2010. Early bursts of body size and shape evolution are
586 rare in comparative data. Evolution 64:2385–2396.
587 Harmon, L. J., J. A. Schulte, A. Larson, and J. B. Losos. 2003. Tempo and mode of
588 evolutionary radiation in iguanian lizards. Science 301:961–964.
589 Harnik, P. G., P. C. Fitzgerald, J. L. Payne, and S. J. Carlson. 2014. Phylogenetic signal in

590 extinction selectivity in Devonian terebratulide brachiopods. Paleobiology 40:675–692.
591 Hill, J. J., M. N. Puttick, T. L. Stubbs, E. J. Rayfield, and P. C. J. Donoghue. 2018. Evolution
592 of jaw disparity in fishes. Palaeontology 61:847–854.
593 Hone, D. W. E., and M. J. Benton. 2005. The evolution of large size: how does Cope's Rule
594 work? Trends Ecol. Evol. 20:4–6.
595 Hopkins, M. J., and S. Gerber. 2017. Morphological disparity. Pp. 1–12 *in* L. N. de la Rosa
596 and G. Muller, eds. Evolutionary developmental biology. Springer, Cham, Switzerland.
597 Hughes, M., S. Gerber, and M. A. Wills. 2013. Clades reach highest morphological disparity
598 early in their evolution. Proc. Natl. Acad. Sci. U. S. A. 110:13875–13879.
599 Korn, D., M. J. Hopkins, and S. A. Walton. 2013. Extinction Space-A Method For The
600 Quantification And Classification Of Changes In Morphospace Across Extinction
601 Boundaries. Evolution 67:2795–2810.
602 Krug, A. Z., and M. E. Patzkowsky. 2015. Phylogenetic clustering of origination and
603 extinction across the late ordovician mass extinction. PLoS One 10:1–11.
604 Lande, R., and S. J. Arnold. 1983. The Measurement of Selection on Correlated Characters.
605 Evolution 37:1210.
606 Liow, L. H., H. J. Skaug, T. Ergon, and T. Schweder. 2010. Global occurrence trajectories of
607 microfossils: environmental volatility and the rise and fall of individual species.
608 Paleobiology 36:224–252.
609 Lockwood, R. 2004. Morphological and ecological patterns of extinction and recovery in
610 veneroid bivalves. 30:507–521.
611 Ma, Y. T. 2010. Birth-death processes on trees. Sci. China Math. 53:2993–3004.
612 McShea, D. W. 1998. Possible largest-scale trends in organismal evolution: Eight "live
613 hypotheses." Annu. Rev. Ecol. Syst. 29:293–318.
614 McShea, D. W., and R. N. Brandon. 2010. Biology's first law: the tendency for diversity and

615 complexity to increase in evolutionary systems. University of Chicago Press.
616 Mooers, A., O. Gascuel, T. Stadler, H. Li, and M. Steel. 2012. Branch lengths on birth-death
617 trees and the expected loss of phylogenetic diversity. Syst. Biol. 61:195–203.
618 Novack-Gottshall, P. M. 2007. Using a theoretical ecospace to quantify the ecological
619 diversity of Paleozoic and Modern marine biotas. 33:273–294.
620 Oyston, J. W., M. Hughes, S. Gerber, and M. A. Wills. 2016. Why should we investigate the
621 morphological disparity of plant clades? Ann. Bot. 117:859–879.
622 Oyston, J. W., M. Hughes, P. J. Wagner, S. Gerber, and M. A. Wills. 2015. What limits the
623 morphological disparity of clades? Interface Focus 5.
624 Pagel, M. 1997. Inferring evolutionary processes from phylogenies. Zool. Scr. 26:331–348.
625 Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.
626 Pie, M. R., and J. S. Weitz. 2005. A null model of morphospace occupation. Am. Nat. 166.
627 Puttick, M. N., T. Ingram, M. Clarke, and G. H. Thomas. 2020. MOTMOT: Models of trait
628 macroevolution on trees (an update). Methods Ecol. Evol. 2020:1–8.
629 Puttick, M. N., J. Kriwet, W. Wen, S. Hu, G. H. Thomas, and M. J. Benton. 2017. Body
630 length of bony fishes was not a selective factor during the biggest mass extinction of all
631 time. Palaeontology 60:727–741.
632 Raup, D. 1981. Extinction: bad genes or bad luck? Acta geológica hispánica 16:25–33.
633 Raup, D., S. J. Gould, T. Schopf, and D. Simberloff. 1973. Models of phylogeny and the
634 evolution of diversity. J. Geol. 81:525–542.
635 Raup, D., and J. Sepkoski. 1982. Mass extinctions in the marine fossil record. Science (80-.).
636 215:1501–1503.
637 Roy, K. 1996. The roles of mass extinction and biotic interaction in large-scale replacements:
638 A reexamination using the fossil record of stromboidean gastropods. Paleiontologische
639 Zeitschrift 22:436–452.

640 Roy, K., and M. Foote. 1997. Morphological approaches to measuring biodiversity. Trends
641 Ecol. Evol. 12:277–281.
642 Ruta, M., K. D. Angielczyk, J. Fröbisch, and M. J. Benton. 2013. Decoupling of
643 morphological disparity and taxic diversity during the adaptive radiation of anomodont
644 therapsids. Proc. R. Soc. B Biol. Sci. 280.
645 Sepkoski, J. J. 1981. A Factor Analytic Description of the Phanerozoic Marine Fossil Record.
646 Paleobiology 7:36–53.
647 Simon, M. S., D. Korn, and S. Koenemann. 2010. Disparity fluctuations in Jurassic
648 ammonoids by means of conch geometry. Palaeogeogr. Palaeoclimatol. Palaeoecol.
649 292:520–531. Elsevier B.V.
650 Simpson, G. 1944. Tempo and mode in evolution. Sci. York 82:036121.
651 Slater, G. J., J. A. Goldbogen, and N. D. Pyenson. 2017. Independent evolution of baleen
652 whale gigantism linked to Plio-Pleistocene ocean dynamics. Proc. R. Soc. B Biol. Sci.
653 284:20170546.
654 Slater, G. J., S. A. Price, F. Santini, and M. E. Alfaro. 2010. Diversity versus disparity and
655 the radiation of modern cetaceans. Proc. R. Soc. - Biol. Sci. 277:3097–3104.
656 Smith, A. 2007. Marine diversity through the Phanerozoic: problems and prospects. J. Geol.
657 Soc. London. 164:731–745.
658 Soul, L. C., and M. Friedman. 2017. Bias in phylogenetic measurements of extinction and a
659 case study of end-Permian tetrapods. Palaeontology 60:169–185.
660 Stadler, T. 2010. Sampling-through-time in birth-death trees. J. Theor. Biol. 267:396–404.
661 Stanley, S. M. 2016. Estimates of the magnitudes of major marine mass extinctions in earth
662 history. Proc. Natl. Acad. Sci. U. S. A. 201613094.
663 Stone, J. R. 1997. The spirit of D'Arcy Thompson dwells in empirical morphospace. Math.
664 Biosci. 142:13–30.

Villier, L., and D. Korn. 2004. Morphological disparity of ammonoids and the mark of Permian mass extinctions. Science 306:264–266.

Wiens, J. J. 2017. What explains patterns of biodiversity across the Tree of Life?: New research is revealing the causes of the dramatic variation in species numbers across branches of the Tree of Life. BioEssays 39:1–10.

Wills, M. A. 1998a. Cambrian and recent disparity: The picture from priapulids. Paleobiology 24:177–199.

Wills, M. A. 1998b. Crustacean disparity through the phanerozoic: Comparing morphological and stratigraphic data. Biol. J. Linn. Soc. 65:455–500.

Wills, M. A. 2001. Morphological disparity: a primer. Pp. 55–144 *in* Fossils, phylogeny, and form. Springer, Boston.

Wills, M. A., D. E. G. Briggs, and R. A. Fortey. 1994. Disparity as an evolutionary index: A comparison of Cambrian and recent arthropods. Paleobiology 20:93–130.

Wills, M. A., S. Gerber, M. Ruta, and M. Hughes. 2012. The disparity of priapulid, archaeopriapulid and palaeoscolecid worms in the light of new data. J. Evol. Biol. 25:2056–2076.

FIGURE LEGENDS

Fig 1. Workflow of simulation and analyses. In the simulations, trees and traits are simulated with time-homogeneous birth-death and Brownian motion processes (A). After 50, 100, 200 contemporary tips are present, a mass extinction removes 50, 75, or 90% of lineages (B). We analysed scenarios in which there is a *directional selectivity* so lineages with larger values above a cut-off are prone to extinction, and *disruptive selectivity* in which trait values closer to the mean are liable to extinction. Strict selectivity means only lineages within the cut-off go extinct, strong selectivity applies the same bias but with a non-zero probability of extinction for all lineages, and in a random extinction all lineages are equally susceptible to extinction. (B). For all simulations disparity is measured for traits using a number of indices (C), with summaries of the levels of diversity and disparity (D).

Fig 2. Summarises of relative and disparity through time in simulations with no mass extinctions for iterations run until 50, 100, and 200 contemporary tips are present. For each simulation time, disparity, and diversity measures are scaled to unity, and each line represents a single simulation. The relative disparity patterns through time from Sum Of Ranges are consistent across iterations, but the patterns from other metrics are more variable.

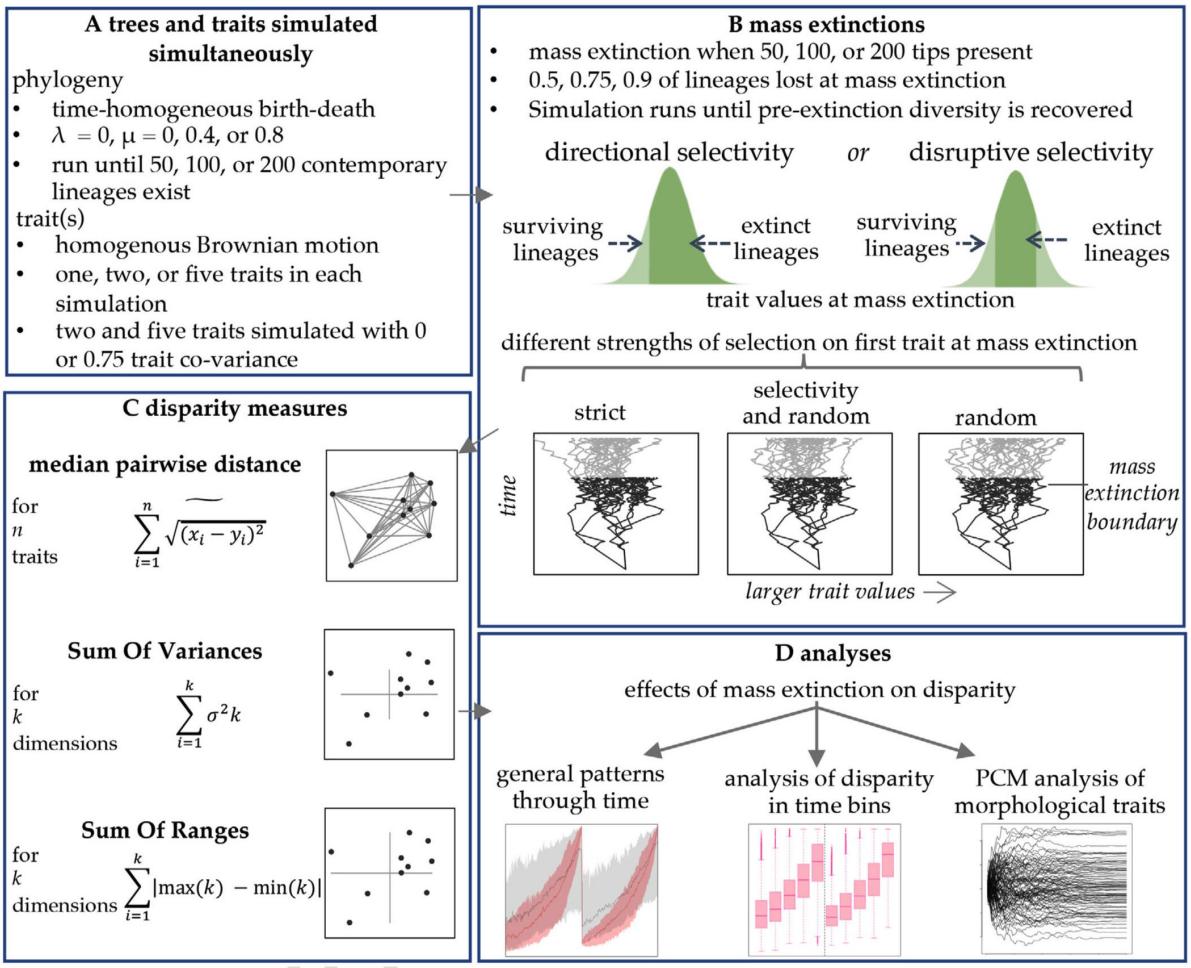
Fig 3. The relative difference between diversity (red lines) and disparity (grey lines) measured in simulations of one trait with 200 extant tips (background extinction 0.8). Results are scaled to show the extinction at the midpoint of each iteration that destroyed 0.5 or 0.9 of contemporary lineages; the median (dark line) and full range (shaded areas) of all iterations are summarised. Full all iterations the variance of disparity indices across all iterations is larger than the trends in diversity patterns. When extinctions are strict or strong diversity and disparity exhibit similar trends with large decreases following extinction, except when disparity is measured using median root distance. Non-selective extinctions do not generally lead to disparity decreases.

Fig 4. Relative change in disparity across a mass extinction boundary for simulations with one (A), two (B), and five (C) traits. The dark horizontal line at zero indicates a null model of no disparity change across an extinction, values below this line indicate a decrease in disparity, and values above indicate a disparity increase. The figures summarise data for all considered disparity indices, including the amount of lineage loss. For two (B) and five (C) traits the first, darker box shows the traits that evolved independently and the lighter, second box shows traits that were simulated with co-variance. Vertical shading denotes selectivity of trait values.

Fig 5. Summary of the number of pGLS (green) and OLS (pink) models that supported a significant relationship between the trait value and extinction for different selectivity levels and severity of extinction. For most simulations there was a true relationship between extinction and trait values (A-C) so it is expected the majority of models would support a significant relationship (gray shading). For the selectivity models (A-C) OLS models more consistently support a significant relationship compared to pGLS, and models perform more poorly as the number of lost lineages at a mass extinction increases. When extinction is random with regards to traits (D) most models should reject a relationship between trait values and extinction (i.e., below 5%, gray shading); OLS models generally have a high type-two error rate but pGLS models consistently and correctly reject a relationship.

Fig 6. The effects of trait distribution and phylogenetic comparative models applied to extant lineages on a tree that when through a deep-time mass extinction. The results are summarised for a directional selectivity (A) in which selectivity is directed towards lineages with larger trait values and disruptive selectivity (B) in which extinction selectivity is directed towards lineages with trait values at the tails of the distribution. All data were simulated under homogeneous Brownian motion, but models applied to extant data only show higher support for the OU model when selectivity is directional (C) and support for the EB model when

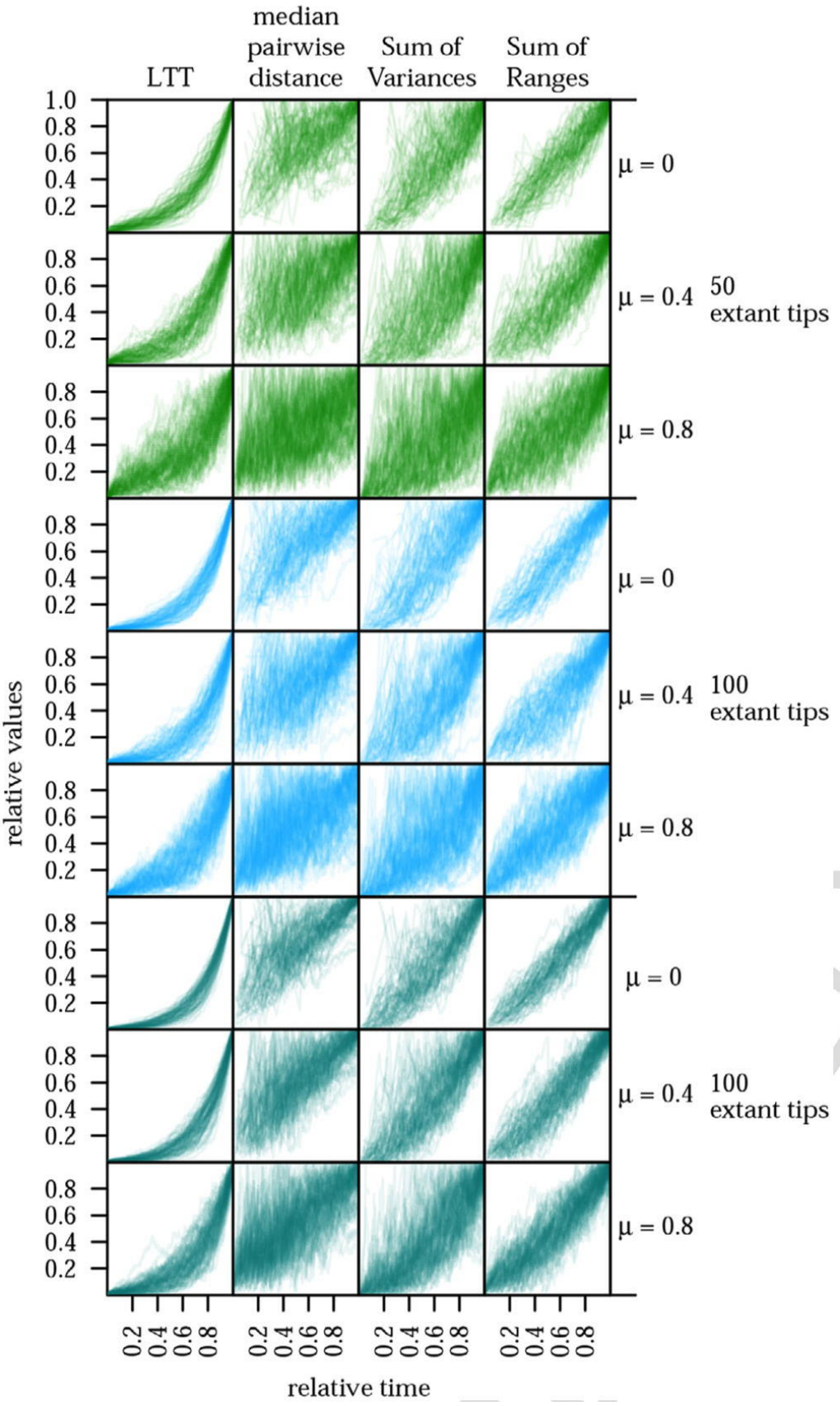
732 selectivity is disruptive (D). A directional selectivity leads to a trait distribution that
733 resembles a distribution expected under an OU process (E), and a disruptive extinction
734 resembles an EB-type distribution (F).
735



736

737 Fig. 1

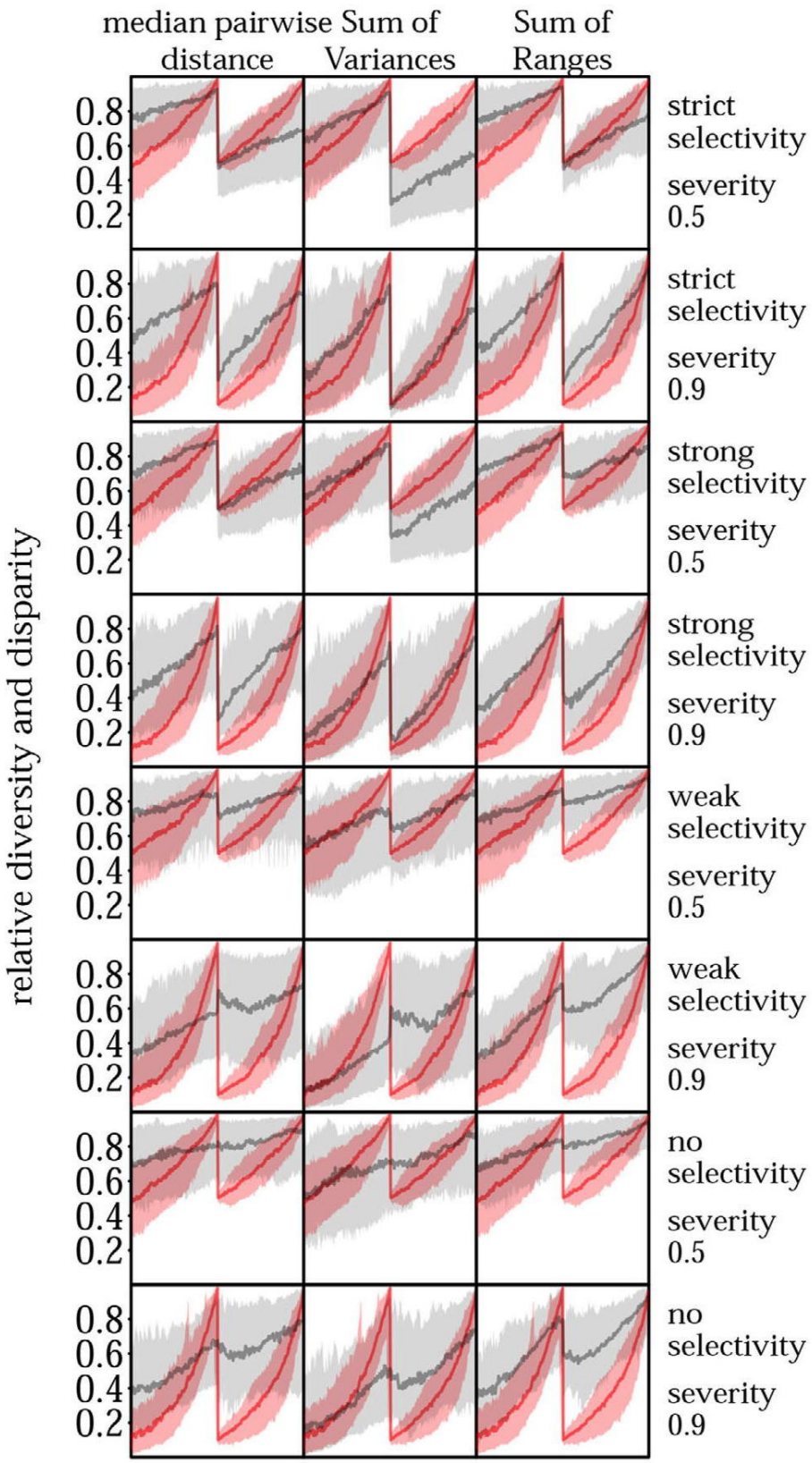
738



739

740 Fig. 2.

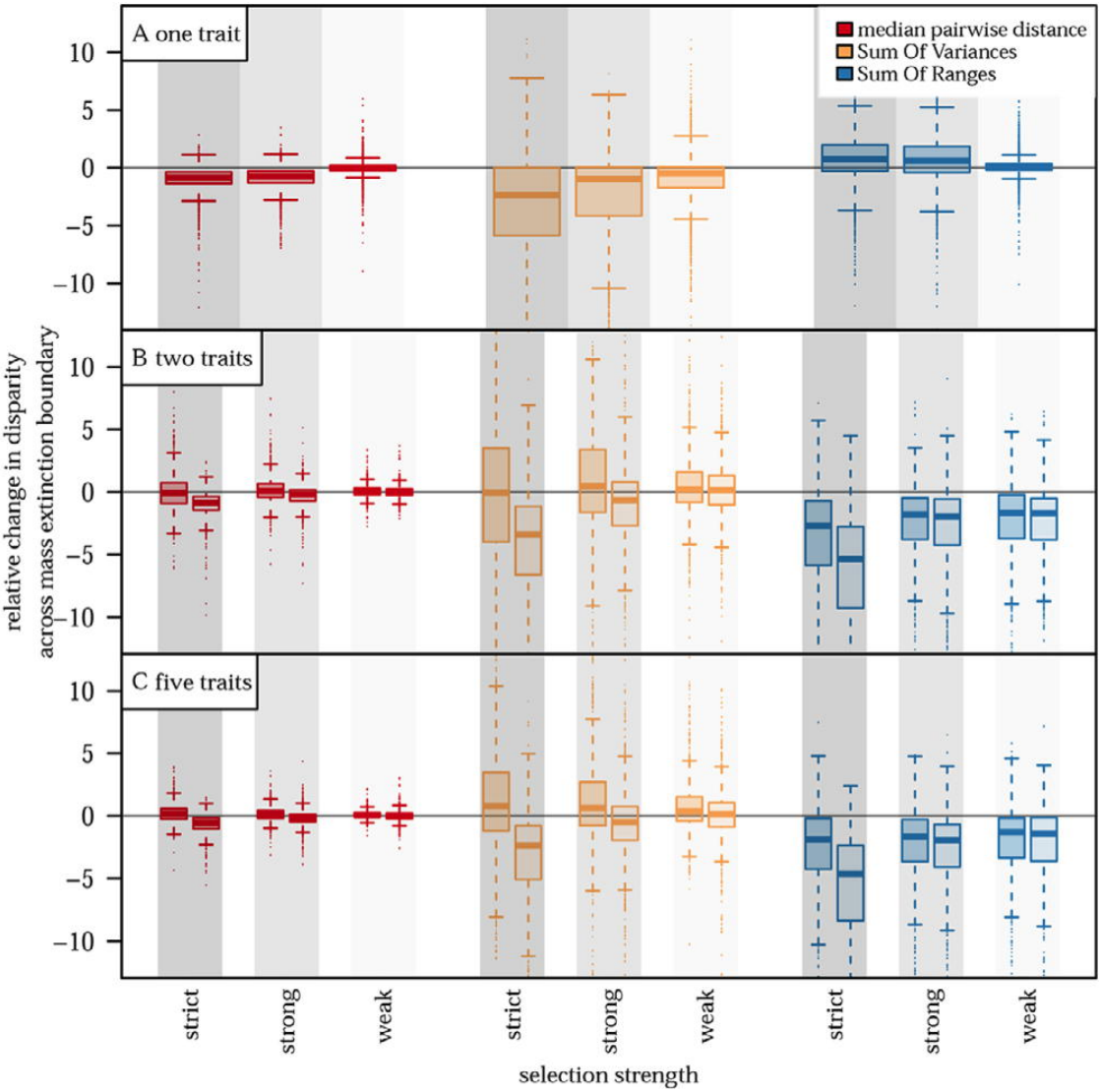
741



742

743 Fig. 3.

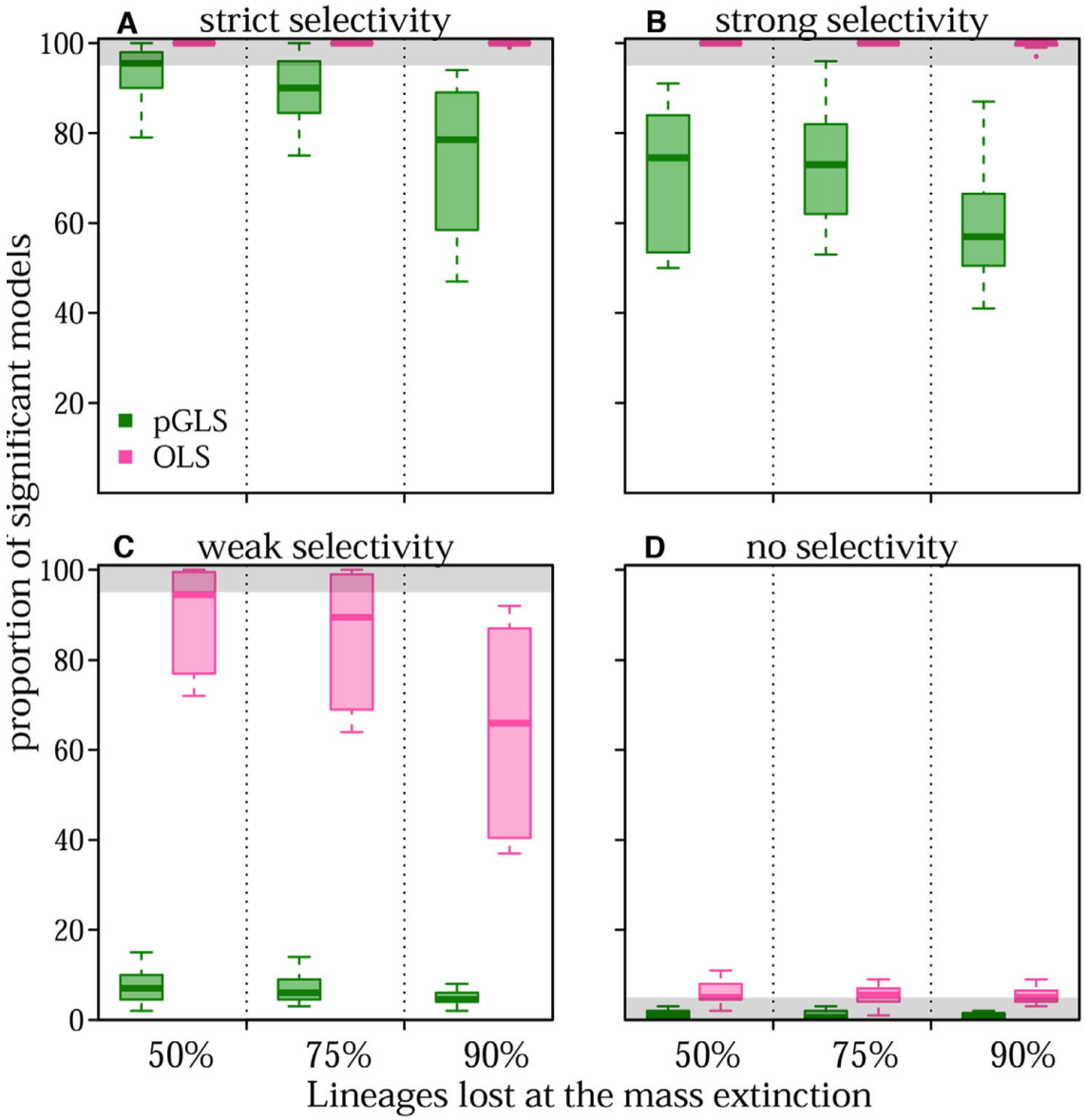
744



745

746 Fig. 4.

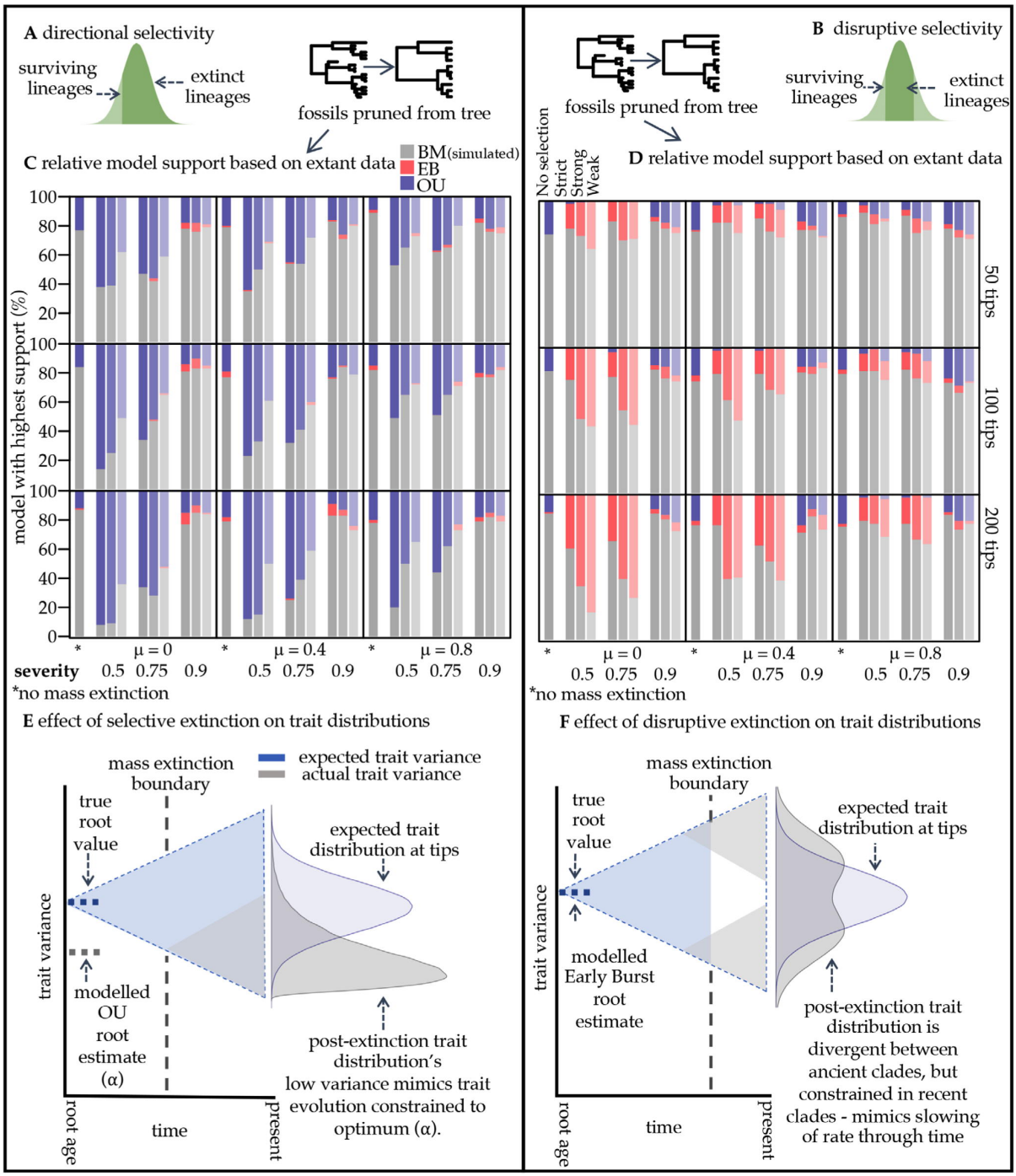
747



748

749 Fig. 5.

750


A directional selectivity
surviving lineages
extinct lineages
fossils pruned from tree
B disruptive selectivity
surviving lineages
extinct lineages
fossils pruned from tree
C relative model support based on extant data
BM(simulated)
EB
OU
model with highest support (%)
D relative model support based on extant data
No selection
Strict
Strong
Weak
50 tips
100 tips
200 tips
μ = 0
μ = 0.4
μ = 0.8
severity 0.5 0.75 0.9
*no mass extinction
E effect of selective extinction on trait distributions
F effect of disruptive extinction on trait distributions
mass extinction boundary
expected trait variance
actual trait variance
true root value
expected trait distribution at tips
trait variance
modelled OU root estimate (α)
post-extinction trait distribution's low variance mimics trait evolution constrained to optimum (α).
modelled Early Burst root estimate
post-extinction trait distribution is divergent between ancient clades, but constrained in recent clades - mimics slowing of rate through time
root age
time
present

751

752 Fig. 6